

# Can Generative Pre-trained Transformers (GPT) Pass Assessments in Higher Education Programming Courses?

Year: 2023 | Citations: 113 | Authors: Jaromir Savelka, Arav Agarwal, C. Bogart, Yifan Song, M. Sakr

---

## Abstract

We evaluated the capability of generative pre-trained transformers (GPT), to pass assessments in introductory and intermediate Python programming courses at the postsecondary level. Discussions of potential uses (e.g., exercise generation, code explanation) and misuses (e.g., cheating) of this emerging technology in programming education have intensified, but to date there has not been a rigorous analysis of the models' capabilities in the realistic context of a full-fledged programming course with diverse set of assessment instruments. We evaluated GPT on three Python courses that employ assessments ranging from simple multiple-choice questions (no code involved) to complex programming projects with code bases distributed into multiple files (599 exercises overall). Further, we studied if and how successfully GPT models leverage feedback provided by an auto-grader. We found that the current models are not capable of passing the full spectrum of assessments typically involved in a Python programming course (<70% on even entry-level modules). Yet, it is clear that a straightforward application of these easily accessible models could enable a learner to obtain a non-trivial portion of the overall available score (>55%) in introductory and intermediate courses alike. While the models exhibit remarkable capabilities, including correcting solutions based on auto-grader's feedback, some limitations exist (e.g., poor handling of exercises requiring complex chains of reasoning steps). These findings can be leveraged by instructors wishing to adapt their assessments so that GPT becomes a valuable assistant for a learner as opposed to an end-to-end solution.